Design of Spectrogram-Consistency Regularization Term Dependent on Observation in Independent Low-Rank Matrix Analysis for Blind Source Separation

Takaaki Kojima*, Norihiro Takamune*, Daichi Kitamura†, and Hiroshi Saruwatari*
* The University of Tokyo, Japan
† National Institute of Technology, Kagawa College, Japan

***Abstract*—Independent low-rank matrix analysis (ILRMA) is the state-of-the-art technique for blind source separation under the overdetermined condition. Recently, some methods that focus on the spectrogram consistency of the separated signals to improve the separation performance have been proposed. One of such methods introduces into ILRMA the regularization term that directly induces the separated signals to be consistent. Although in the conventional method, the regularization term is designed to be independent of the observed signals, we design the regularization term to be dependent on the observed signals. For the obtained cost function, we derive a new update rule on the basis of the majorization-minimization algorithm with the auxiliary function that is an extension of that derived in the conventional method. Finally, a numerical experiment is conducted to verify the separation performance.**

I. Introduction

Blind source separation (BSS) is a technique for estimating individual sources from an observed mixture without knowledge of the characteristics of sources or mixing conditions. Independent component analysis (ICA) [1] is a representative approach to solving BSS under the overdetermined condition (the number of microphones M is greater than or equal to that of sources N). In ICA, the demixing matrix is estimated under the assumption that individual sources are statistically independent and the mixing system is instantaneous. Since the sources are mixed by time-domain convolution for acoustic signals, ICA is often applied in the time-frequency domain via short-time Fourier transform (STFT) under the assumption that the mixing system can be approximated by instantaneous mixing in the time-frequency domain. This method is called frequency-domain ICA (FDICA) [2] and estimates frequency-wise demixing matrices by independently applying ICA to complex-valued signals in each frequency bin. ICA has arbitrariness to the permutation of the estimated signals; therefore, the permutations should be aligned for frequency bins in FDICA (permutation problem). Several permutation solvers have been studied to solve this problem [3]–[5]. In contrast, independent vector analysis (IVA) [6] and independent low-rank matrix analysis (ILRMA) [7] can estimate the frequency-wise permutations simultaneously with the demixing matrix. These techniques assume a more complex source model than FDICA and can therefore capture dependences between frequency bins and align the frequency-wise permutations. Among these methods, it has been reported that ILRMA tends to show a higher separation performance than FDICA and IVA [7]. However, it has been reported that there is room for performance improvement even in ILRMA [8].

The spectrogram consistency (or simply *consistency*) [9]–[11] has been studied as a new promising clue to align the frequency-wise permutations [12]–[14]. In some BSS methods in the time-frequency domain, such as FDICA, IVA, and ILRMA, assumptions on the spectrograms of the separated signals are made. However, these assumptions are not guaranteed to be satisfied for the final output in the time domain obtained as the inverse STFT (ISTFT) of the spectrograms of the separated signals, i.e., the spectrograms obtained as the STFT of the final output are often different from the spectrograms of the separated signals before ISTFT. This is because the operation of STFT after ISTFT is a projection operator on the image space of the STFT as a linear operator. The image space of the STFT is called the consistent subspace in this paper. For example, if the frequency-wise permutations are not aligned, the spectrogram is changed greatly by the projection and becomes blurred in both time and frequency directions as shown in Fig. 1. Then, the performance of BSS is expected to be improved by making assumptions on the projected spectrograms or by reducing the difference between the spectrograms before and after the projection. It has been reported that the performance of FDICA without a permutation solver can be improved by incorporating the projection operator into the cost function of FDICA [12].

As for ILRMA with the spectrogram consistency considered, consistent ILRMA (Consist. ILRMA) [13] and ILRMA with spectrogram-consistency regularization [14] have been proposed. In Consist. ILRMA [13], the update rules of ILRMA are modified to approximate the projected spectrograms instead of the spectrograms of the separated signals (before the

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011 (for algorithm development) and Tateisi Science and Technology Foundation (for practical experiment).

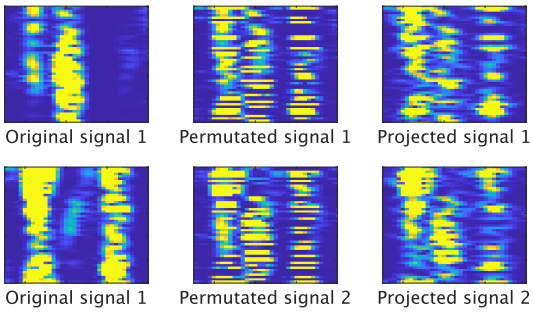


Fig. 1: Examples of the spread by projecting inconsistent spectrograms onto the consistent subspace. Original speech signals (left) were permutated randomly (middle) with frequency bins and projected onto the consistent subspace (right).

projection) by nonnegative matrix factorization (NMF) [15]. In [14], a regularization term that induces the spectrograms of the separated signals (before the projection) themselves to be close to the consistent subspace is designed for ILRMA. ILRMA with the spectrogram-consistency regularization has the advantage of being formulated as an optimization problem, whereas Consist. ILRMA is heuristic and unfortunately not formulated as an optimization problem. The regularization term proposed in [14] is designed to be independent of the observed signals; in other words, it is determined solely by the STFT conditions (e.g., the window length, window function, and hopsize). Therefore, neither the specific structure nor correlation of the observed signals can be fully utilized in [14]. In this paper, we design a new regularization term for ILRMA on the basis of the distances between the spectrograms of the separated signals and the projected spectrograms of the separated signals onto the consistent subspace. The proposed regularization term depends on the observed signals; thus, the separation performance is expected to be improved more than in [14]. For the new cost function, we unfortunately cannot use the same technique of [14] to design the auxiliary function for the majorization-minimization (MM) algorithm [16]. To address this problem, we extend the auxiliary-function-design methodology from [14] by utilizing the mathematical finding in this paper and design an auxiliary function on the basis of this finding. We derive a new update rule by applying vector-wise coordinate descent (VCD) [17] to the designed auxiliary function. Finally, we verify the separation performance of the proposed method by a simulation experiment.

II. PRELIMINARIES

In this paper, we consider a time-domain signal to be a real-valued signal of length L. The complex-valued spectrogram of a time-domain signal $\tilde{\boldsymbol{\psi}} \in \mathbb{R}^L$ is represented by the supervecter $\vec{\boldsymbol{\Psi}} \in \mathbb{C}^{IJ}$, where I and J are the window length and the number of time frames in STFT, respectively. Here, $(I(j-1)+i)$th element of $\vec{\boldsymbol{\Psi}}$ is the ith frequency bin of the spectrum in the jth frame, where $i = 1, \ldots, I$ and $j = 1, \ldots, J$ are the indices of the frequency bins and time frames, respectively. Hereafter, we assume the practical case where $IJ > L$ holds. Let τ be the hopsize of STFT, $\boldsymbol{h} \in \mathbb{R}^I$ be the window function. The length of the signal after zero-padding is denoted by $L' = L^{(\mathrm{pre})} + L + L^{(\mathrm{post})}$, where $L^{(\mathrm{pre})} (< I)$ and $L^{(\mathrm{post})} (< I)$ are the length of zero-padding at the beginning and end of the time signal, respectively. Note that $L' = \tau(J-1) + I$ holds. The $L' \times L$ matrix representing zero-padding is given by

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{O}_{L \times L^{(\mathrm{pre})}} & \boldsymbol{E}_L & \boldsymbol{O}_{L \times L^{(\mathrm{post})}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{L' \times L}, \quad (1)$$

where $\boldsymbol{O}_{q \times q'}$ is the $q \times q'$ zero matrix and $\boldsymbol{E}_q$ is the $q \times q$ identity matrix for any nonnegative integer q, q'. Here, $\cdot^{\mathsf{T}}$ represents the transpose. That is, the signal after zero padding can be represented by $\boldsymbol{Z}\tilde{\boldsymbol{\psi}}$. The $(I(j-1)+1)$th to (Ij)th elements of $\vec{\boldsymbol{\psi}}$, which is the spectrum in the jth frame, are obtained by multiplying the $(\tau(j-1)+1)$th to $(\tau(j-1)+I)$th samples of the zero-padded signal $\boldsymbol{Z}\tilde{\boldsymbol{\psi}}$ by the window function $\boldsymbol{h}$ and applying discrete Fourier transform (DFT). Thus, the vector created by arranging the $(I(j-1)+1)$th to (Ij)th element of $\vec{\boldsymbol{X}}$ is represented by $\boldsymbol{F}\boldsymbol{\Delta}_j\boldsymbol{Z}\tilde{\boldsymbol{\psi}}$, where $\boldsymbol{F} \in \mathbb{C}^{I \times I}$ denotes the representation matrix of DFT and $\boldsymbol{\Delta}_j \in \mathbb{R}^{I \times L'}$ is defined as

$$\boldsymbol{\Delta}_j = \begin{bmatrix} \boldsymbol{O}_{I \times \tau(j-1)} & \mathrm{diag}\{\boldsymbol{h}\} & \boldsymbol{O}_{I \times (L' - \tau(j-1) - I)} \end{bmatrix}. \quad (2)$$

Here, $\mathrm{diag}\{\boldsymbol{h}\}$ represents the diagonal matrix whose diagonal elements are $\boldsymbol{h}$. Therefore, $\vec{\boldsymbol{\Psi}}$ is represented by $\vec{\boldsymbol{\Psi}} = \mathcal{F}\tilde{\boldsymbol{\psi}}$, where $\mathcal{F} \in \mathbb{C}^{IJ \times L}$ is defined as

$$\mathcal{F} = \begin{bmatrix} \boldsymbol{F}\boldsymbol{\Delta}_1\boldsymbol{Z} \\ \vdots \\ \boldsymbol{F}\boldsymbol{\Delta}_J\boldsymbol{Z} \end{bmatrix}. \quad (3)$$

In this paper, we consider that the STFT conditions satisfy the condition that $\mathcal{F}$ has full column rank so that ISTFT exists. We use the Moore–Penrose inverse $\cdot^{\dagger}$ of $\mathcal{F}$ as ISTFT, i.e., $\mathcal{F}^{\dagger}$ is a representation matrix of ISTFT. Note that $\mathcal{F}^{\dagger}\mathcal{F}$ is equal to $\boldsymbol{E}_L$ and $\mathcal{F}\mathcal{F}^{\dagger}$ is an orthogonal projection matrix onto the image space of $\mathcal{F}$. A spctrogram $\vec{\boldsymbol{\Psi}}' \in \mathbb{C}^{IJ}$ is called *consistent* when $\vec{\boldsymbol{\Psi}}'$ is in the image space of $\mathcal{F}$, i.e., $\vec{\boldsymbol{\Psi}}' = \mathcal{F}\mathcal{F}^{\dagger}\vec{\boldsymbol{\Psi}}'$ holds.

III. CONVENTIONAL METHODS

A. ILRMA[7]

Let $s_{ij,n}$, $x_{ij,m}$, and $y_{ij,n}$ be the (i,j)th bin in the spectrogram of the nth source, the mth observed, and nth separated signals, respectively, where $n = 1, \ldots, N$ and $m = 1, \ldots, M$ denote the indices of the source and observed signals, respectively. We define $\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^{\mathsf{T}} \in \mathbb{C}^N$, $\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^{\mathsf{T}} \in \mathbb{C}^M$, and $\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^{\mathsf{T}} \in \mathbb{C}^N$.When the mixing system is time-invariant and the reverberation time is sufficiently shorter than the length of the STFT window, we can assume instantaneous mixing in the time-frequency domain as $\boldsymbol{x}_{ij} = \boldsymbol{A}_i\boldsymbol{s}_{ij}$, where $\boldsymbol{A}_i \in \mathbb{C}^{M \times N}$ is the mixing matrix. In this paper, we assume that $M = N$ and $\boldsymbol{A}_i$ is regular. Under these assumptions, there exists the demixing matrix $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,N})^{\mathsf{H}} \simeq \boldsymbol{A}_i^{-1}$ that satisfies $\boldsymbol{y}_{ij} = \boldsymbol{W}_i\boldsymbol{x}_{ij}$. Let $w_{i,nm}$ be the mth elements of $\boldsymbol{w}_{i,n}$.

In ILRMA, $y_{ij,n}$ is assumed to be generated from a complex Gaussian distribution with a mean of zero and the variance $\rho_{ij,n}$ as follows:

$$p(y_{ij,n}) = \frac{1}{\pi\rho_{ij,n}} \exp\left(-\frac{|y_{ij,n}|^2}{\rho_{ij,n}}\right). \tag{4}$$

The variance $\rho_{ij,n}$ is assumed to have a low-rank structure and modeled by NMF [15] as

$$\rho_{ij,n} = \sum_{k=1}^{K} t_{ik,n} v_{kj,n}, \tag{5}$$

where $t_{ik,n} \geq 0$ and $v_{kj,n} \geq 0$ represent the basis and the activation, respectively, and $k = 1, \ldots, K$ denotes the index of the basis. The demixing matrix $\{\boldsymbol{W}_i\}$ and the NMF variables $\{t_{ik,n}\}$, $\{v_{kj,n}\}$ are obtained from maximum likelihood estimation by minimizing the following negative log-likelihood as a cost function:

$$\begin{aligned}\mathcal{L} =& \frac{1}{2}\sum_{i,j,n}\left(\frac{|\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_k t_{ik,n}v_{kj,n}} + \log\sum_k t_{ik,n}v_{kj,n}\right) \\ & - \frac{J}{2}\sum_i \log|\det \boldsymbol{W}_i|^2 + \text{const.},\end{aligned} \tag{6}$$

where const. denotes the terms independent of $\{\boldsymbol{W}_i\}$, $\{t_{ik,n}\}$ and $\{v_{kj,n}\}$. To optimize (6), the demixing matrix $\{\boldsymbol{W}_i\}$ and the NMF variables $\{t_{ik,n}\}$, $\{v_{kj,n}\}$ are updated alternately. The demixing matrix $\{\boldsymbol{W}_i\}$ is updated with iterative projection (IP) [18]. The NMF variables $\{t_{ik,n}\}$, $\{v_{kj,n}\}$ are updated on the basis of the MM algorithm[16]. All of these update rules ensure the monotonic nonincrease in the cost function (6).

B. ILRMA with spectrogram-consistency regularization independent of observation [14]

In [14], the regularization term based on spectrogram consistency is introduced to ILRMA. Let $\vec{\boldsymbol{Y}}_n$ be the supervectorized spectrogram of the nth separated signal whose $(I(j-1)+i)$th element is $y_{ij,n}$, and the degree to which $\vec{\boldsymbol{Y}}_n$ deviates from the consistent subspace is formulated as [14]

$$\begin{aligned}\mathcal{E}(\vec{\boldsymbol{Y}}_n) &= \|\vec{\boldsymbol{Y}}_n - \mathcal{F}\mathcal{F}^{\dagger}\vec{\boldsymbol{Y}}_n\|_2^2 \\ &= \left\|(\boldsymbol{E}_{IJ} - \mathcal{F}\mathcal{F}^{\dagger})\sum_m \mathcal{W}_{nm}\mathcal{F}\tilde{\boldsymbol{x}}_m\right\|_2^2,\end{aligned} \tag{7}$$

where $\mathcal{W}_{nm} \in \mathbb{C}^{IJ\times IJ}$ is a diagonal matrix whose $(I(j-1)+i)$th diagonal element is $w_{i,nm}^*$ and $\tilde{\boldsymbol{x}}_m \in \mathbb{R}^L$ is the mth observed signal in the time domain. Here, $\cdot^*$ denotes the complex conjugate. To induce consistency for arbitrary observed signals, the case where $\tilde{\boldsymbol{x}}_m$ is a white Gaussian signal without correlation between channels is considered, and the expected value $\sum_n \mathbb{E}[\mathcal{E}(\vec{\boldsymbol{Y}}_n)]$ is used as a regularization term in [14]. In this paper, the regularization used in [14] is called observation-independent consistency regularization (OICR).

The regularization term can be represented as the quadratic form of $\boldsymbol{w}_{i,n}$ as follows:

$$\sum_n \mathbb{E}\left[\mathcal{E}(\vec{\boldsymbol{Y}}_n)\right] = \sum_{n,i,i'} \boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{ii'}^{(\mathrm{ind})}\boldsymbol{w}_{i',n}, \tag{8}$$

where $\boldsymbol{G}_{ii'}^{(\mathrm{ind})} \in \mathbb{C}^{M\times M}$ is only determined by the STFT conditions. Since $\tilde{\boldsymbol{x}}_m$ is assumed to be independent and identically distributed with respect to m, $\boldsymbol{G}_{ii'}^{(\mathrm{ind})} = g_{ii'}^{(\mathrm{ind})}\boldsymbol{E}_M$ $(g_{ii'}^{(\mathrm{ind})} \in \mathbb{C})$ holds. The cost function of ILRMA with OICR is formulated as follows:

$$\mathcal{L}_{\mathrm{OICR}} = \mathcal{L} + \beta \sum_{n,i,i'} g_{ii'}^{(\mathrm{ind})}\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{w}_{i'n}, \tag{9}$$

where $\beta > 0$ is the weight of the regularization term in ILRMA with OICR. Since the added regularization term (8) is independent of the NMF variables, the update rules for the NMF variables $\{t_{ik,n}\}$ and $\{v_{kj,n}\}$ are the same as in ILRMA. For $\{\boldsymbol{W}_i\}$, an auxiliary function is designed for the MM algorithm [16] and updated using VCD [17] instead of IP (see details in [14]). The update rule of $\{\boldsymbol{W}_i\}$ also guarantees the monotonic nonincrease in the cost function (9).

IV. Proposed method

A. Motivation and approach

Since the regularization term in [14] is formulated to be independent of the to-be-separated observed signals, the regularization cannot be appropriate for specific observed signals. Therefore, if we can introduce a regularization term that depends on the specific structure of the to-be-separated observed signals, further improvement in separation performance is expected. In this paper, we propose the use of $\sum_n \mathcal{E}(\vec{\boldsymbol{Y}}_n)$ itself calculated with the to-be-separated observed signals as a regularization term for ILRMA. The regularization term in the conventional method has a simple structure because there is no correlation between the observed channels. Unfortunately, the proposed regularization term does not always have a simple structure because there exists a correlation between the observed channels. Thus, we cannot exactly follow the technique for designing the auxiliary function as in [14] for the newly obtained optimization problem. We find out the relationship between two matrices based on the positive semidefiniteness that allows us to extend the technique in [14], and we design an auxiliary function for the proposed cost function. On the basis of the MM algorithm, we derive an update rule by applying VCD [17] to the proposed auxiliary function.

B. Design of regularization term and update rules

By expanding (7), the proposed regularization term is also deformed into the following quadratic form of $\boldsymbol{w}_{i,n}$:

$$\sum_n \mathcal{E}(\vec{\boldsymbol{Y}}_n) = \sum_{n,i,i'} \boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{ii'}\boldsymbol{w}_{i',n}, \tag{10}$$

where $\boldsymbol{G}_{ii'} \in \mathbb{C}^{M\times M}$ is dependent on not only the STFT conditions but also the to-be-separated observed signals. The specific representation of $\boldsymbol{G}_{ii'}$ will be omitted for space limitation. The cost function of the proposed method is the weighted sum of the cost function of ILRMA (6) and the regularization term (10) as follows:

$$\mathcal{L}_{\mathrm{Prop}} = \mathcal{L} + \tilde{\gamma}\sum_{n,i,i'} \boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{ii'}\boldsymbol{w}_{i',n}, \tag{11}$$

where $\tilde{\gamma} \geq 0$ is the weight of the regularization term in the proposed method. Then, we derive update rules for the cost function (11). Since the regularization term (10) is independent of the NMF variables $\{t_{ik,n}\}$ and $\{v_{kj,n}\}$ as in [14], the update rules for the NMF variables are the same as in ILRMA. Therefore, it is sufficient to derive an update rule for $\{\boldsymbol{w}_{i,n}\}$.

The cost function (11) has a correlation term between frequencies and this makes it difficult to optimize $\{\boldsymbol{w}_{i,n}\}$ simultaneously. Therefore, we focus on one frequency bin i and treat the variables related to the other frequency bins as constants. That is, $\{\boldsymbol{w}_{i,n}\}$ is updated one by one for each frequency bin in a block coordinate descent manner. The cost function is rewritten with regard to the frequency i as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{Prop}} =& J\left[\sum_n \boldsymbol{w}_{i,n}^{\mathsf{H}} \frac{1}{J}\sum_j \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{\sum_k t_{ik,n}v_{kj,n}}\boldsymbol{w}_{i,n} - \log|\det \boldsymbol{W}_i|^2\right] \\ &+\tilde{\gamma}\sum_n \begin{bmatrix}\boldsymbol{w}_{i,n}\\ \boldsymbol{w}_{i,n}^*\end{bmatrix}^{\mathsf{H}} \begin{bmatrix}\boldsymbol{G}_{ii} & \boldsymbol{G}_{i\hat{i}}\\ \boldsymbol{G}_{\hat{i}i} & \boldsymbol{G}_{\hat{i}\hat{i}}\end{bmatrix}\begin{bmatrix}\boldsymbol{w}_{i,n}\\ \boldsymbol{w}_{i,n}^*\end{bmatrix} \\ &+2\tilde{\gamma}\sum_n \mathrm{Re}\left(\begin{bmatrix}\sum_{i'\neq i,\hat{i}} \boldsymbol{G}_{i'i}^{\mathsf{H}}\boldsymbol{w}_{i',n}\\ \sum_{i'\neq i,\hat{i}} \boldsymbol{G}_{i'\hat{i}}^{\mathsf{H}}\boldsymbol{w}_{i',n}\end{bmatrix}^{\mathsf{H}}\begin{bmatrix}\boldsymbol{w}_{i,n}\\ \boldsymbol{w}_{i,n}^*\end{bmatrix}\right) + \mathrm{const.},\end{aligned} \tag{12}$$

where $\mathrm{Re}\,(\cdot)$ denotes a real part of a complex number and const. denotes the terms independent of $\boldsymbol{w}_{i,n}$. The index $\hat{i}$ is the frequency index opposite of i across the Nyquist frequency, where $i + \hat{i} = I + 2$ $(i \neq 1)$. It is difficult to optimize (12) analytically because the terms $\boldsymbol{w}_{i,n}^{\mathsf{T}}\boldsymbol{G}_{\hat{i}i}\boldsymbol{w}_{i,n}$ and $\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{i\hat{i}}\boldsymbol{w}_{i,n}^*$ are included in the cost function. Note that this problem does not occur in the case $i = 1$ (corresponding to 0 Hz) or $i = I/2$, because $\boldsymbol{w}_{i,n}$ $(i = 1, I/2)$ is real-valued to satisfy the constraint that the separated signal is real-valued in the time domain. The demixing matrix related to $i = 1, I/2$ can be updated using VCD without auxiliary functions. We design the auxiliary function without the terms $\boldsymbol{w}_{i,n}^{\mathsf{T}}\boldsymbol{G}_{\hat{i}i}\boldsymbol{w}_{i,n}$ and $\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{i\hat{i}}\boldsymbol{w}_{i,n}^*$ in the case $i \neq 1, I/2$.

In this paper, we follow the strategy used in [14] to derive the auxiliary function on the basis of Claim IV.1 (Claim 1 in [14]) for (12).

Claim IV.1 (Claim 1 in [14])**.** For any integer $q \geq 1$, $\boldsymbol{w} \in \mathbb{C}^q$, $\boldsymbol{\alpha} \in \mathbb{C}^q$, and positive semidefinite matrix $\hat{\boldsymbol{G}} \in \mathbb{C}^{q\times q}$, we have

$$\begin{aligned}\boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{G}}\boldsymbol{w} \leq& \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{G}}^{+}\boldsymbol{w} \\ &-2\mathrm{Re}\left(\boldsymbol{\alpha}^{\mathsf{H}}(\hat{\boldsymbol{G}}^{+}-\hat{\boldsymbol{G}})\boldsymbol{w}\right)+\boldsymbol{\alpha}^{\mathsf{H}}(\hat{\boldsymbol{G}}^{+}-\hat{\boldsymbol{G}}))\boldsymbol{\alpha},\end{aligned} \tag{13}$$

where $\hat{\boldsymbol{G}}^{+}$ is a positive semidefinite matrix satisfying $\hat{\boldsymbol{G}}^{+} - \hat{\boldsymbol{G}} \succeq \boldsymbol{O}$. Equality holds when $\boldsymbol{\alpha} = \boldsymbol{w}$.

Let $\hat{\boldsymbol{G}}^{+}$, $\hat{\boldsymbol{G}}$, $\boldsymbol{w}$, and $\boldsymbol{\alpha}$ in Claim IV.1 be set as

$$\hat{\boldsymbol{G}}^{+} = \begin{bmatrix}\boldsymbol{G}_{ii}^{+} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{G}_{\hat{i}\hat{i}}^{+},\end{bmatrix}, \hat{\boldsymbol{G}} = \begin{bmatrix}\boldsymbol{G}_{ii} & \boldsymbol{G}_{i\hat{i}}\\ \boldsymbol{G}_{\hat{i}i} & \boldsymbol{G}_{\hat{i}\hat{i}}\end{bmatrix},$$

$\boldsymbol{w} = [\boldsymbol{w}_{i,n}^{\mathsf{T}}, \boldsymbol{w}_{i,n}^{*\mathsf{T}}]^{\mathsf{T}}$, and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{i,n}^{\mathsf{T}}, \boldsymbol{\alpha}_{i,n}^{*\mathsf{T}}]^{\mathsf{T}}$, where $\boldsymbol{G}_{ii}^{+}$ and $\boldsymbol{G}_{\hat{i}\hat{i}}^{+}$ are chosen so that $\hat{\boldsymbol{G}}^{+} - \hat{\boldsymbol{G}} \succeq \boldsymbol{O}$ is satisfied. By using (13), we can derive an auxiliary function without the terms $\boldsymbol{w}_{i,n}^{\mathsf{T}}\boldsymbol{G}_{\hat{i}i}\boldsymbol{w}_{i,n}$ and $\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{G}_{i\hat{i}}\boldsymbol{w}_{i,n}^*$ for the second term of (12).

In [14], since $\boldsymbol{G}_{ii'}^{(\mathrm{ind})}$ becomes the diagonal matrix due to no correlation with respect to m, the problem of finding the matrix $\hat{\boldsymbol{G}}^{+}$ is simplified to that of finding 2×2 matrices, and the following relation is used in [14]:

$$\begin{bmatrix}g_{ii}^{(\mathrm{ind})}+|g_{i\hat{i}}^{(\mathrm{ind})}| & 0\\ 0 & g_{\hat{i}\hat{i}}^{(\mathrm{ind})}+|g_{\hat{i}i}^{(\mathrm{ind})}|\end{bmatrix} - \begin{bmatrix}g_{ii}^{(\mathrm{ind})} & g_{i\hat{i}}^{(\mathrm{ind})}\\ g_{\hat{i}i}^{(\mathrm{ind})} & g_{\hat{i}\hat{i}}^{(\mathrm{ind})}\end{bmatrix} \succeq \boldsymbol{O}. \tag{14}$$

Since $\boldsymbol{G}_{ii'}$ has nondiagonal terms due to the correlation regarding m, we cannot, however, simply follow this derivation in [14]. For this problem, we find the following relation as

$$\begin{bmatrix}\boldsymbol{G}_{ii}+(\boldsymbol{G}_{i\hat{i}}\boldsymbol{G}_{\hat{i}i})^{\frac{1}{2}} & \boldsymbol{O}_{M\times M}\\ \boldsymbol{O}_{M\times M} & \boldsymbol{G}_{\hat{i}i}+(\boldsymbol{G}_{\hat{i}i}\boldsymbol{G}_{i\hat{i}})^{\frac{1}{2}}\end{bmatrix} - \begin{bmatrix}\boldsymbol{G}_{ii} & \boldsymbol{G}_{i\hat{i}}\\ \boldsymbol{G}_{\hat{i}i} & \boldsymbol{G}_{\hat{i}\hat{i}}\end{bmatrix} \succeq \boldsymbol{O}, \tag{15}$$

which is an extension of (14). This fact can be proved by singular value decomposition, but we omit the details for space limitation.

By using Claim IV.1 and (15), we design the total auxiliary function for (12) as

$$\begin{aligned}\mathcal{L}_{\mathrm{Prop}}^{+} =& J\left[\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{D}_{i,n}^{+}\boldsymbol{w}_{i,n} + 2\mathrm{Re}\,(\boldsymbol{b}_{i,n}^{+\mathsf{H}}\boldsymbol{w}_{i,n}) - \log|\det \boldsymbol{W}_i|^2\right] \\ &+ \mathrm{const.},\end{aligned} \tag{16}$$

where const. denotes the terms independent of $\boldsymbol{w}_{i,n}$ and $\boldsymbol{D}_{i,n}^{+}$ and $\boldsymbol{b}_{i,n}^{+}$ are respectively defined as

$$\boldsymbol{D}_{i,n}^{+} = \frac{1}{J}\sum_j \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{\sum_k t_{ik,n}v_{kj,n}} + \frac{2\tilde{\gamma}}{J}\left(\boldsymbol{G}_{ii}+(\boldsymbol{G}_{i\hat{i}}\boldsymbol{G}_{\hat{i}i})^{\frac{1}{2}}\right), \tag{17}$$

$$\boldsymbol{b}_{i,n}^{+} = \frac{2\tilde{\gamma}}{J}\left[\sum_{i'\neq i,i}\boldsymbol{G}_{i'i}^{\mathsf{H}}\boldsymbol{w}_{i',n} - (\boldsymbol{G}_{i\hat{i}}\boldsymbol{G}_{\hat{i}i})^{\frac{1}{2}}\boldsymbol{\alpha}_{i,n} + \boldsymbol{G}_{\hat{i}i}^{\mathsf{H}}\boldsymbol{\alpha}_{i,n}^{*}\right]. \tag{18}$$

Here, $\boldsymbol{\alpha}_{i,n} \in \mathbb{C}^M$ is an auxiliary variable and the equality for the auxiliary function holds when $\boldsymbol{\alpha}_{i,n} = \boldsymbol{w}_{i,n}$. Since (16) is the sum of a quadratic form, a linear term, and a log-determinant term, an update rule using VCD [17] can be obtained as follows:

$$\boldsymbol{\zeta}_{i,n} \leftarrow (\boldsymbol{W}_i\boldsymbol{D}_{i,n}^{+})^{-1}\boldsymbol{e}_n, \tag{19}$$

$$\hat{\boldsymbol{\zeta}}_{i,n} \leftarrow (\boldsymbol{D}_{i,n}^{+})^{-1}\boldsymbol{b}_{i,n}^{+}, \tag{20}$$

$$\chi_{i,n} \leftarrow \boldsymbol{\zeta}_{i,n}^{\mathsf{H}}\boldsymbol{D}_{i,n}^{+}\boldsymbol{\zeta}_{i,n}, \tag{21}$$

$$\hat{\chi}_{i,n} \leftarrow \boldsymbol{\zeta}_{i,n}^{\mathsf{H}}\boldsymbol{D}_{i,n}^{+}\hat{\boldsymbol{\zeta}}_{i,n}, \tag{22}$$

$$\boldsymbol{w}_{i,n} \leftarrow \begin{cases}\frac{\boldsymbol{\zeta}_{i,n}}{\sqrt{\chi_{i,n}}} - \hat{\boldsymbol{\zeta}}_{i,n} & (\hat{\chi}_{i,n} = 0),\\ \frac{\hat{\chi}_{i,n}}{2\chi_{i,n}}\left(1-\sqrt{1+\frac{4\chi_{i,n}}{|\hat{\chi}_{i,n}|^2}}\right)\boldsymbol{\zeta}_{i,n} - \hat{\boldsymbol{\zeta}}_{i,n} & (\text{otherwise}),\end{cases} \tag{23}$$

where $\boldsymbol{e}_n \in \mathbb{C}^N$ is a vector whose nth component is one and the others are zero. This update rule guarantees the monotonic nonincrease in the cost function (11). All the update rules in

TABLE I: Averages of SDR improvements over 100 trials for each method

ILRMA	Consist. ILRMA	ILRMA w/ OSICR			Prop. w/o temp.			Prop. w/ temp.		
		$\beta = 10^{-6}$	$\beta = 10^{-5}$	$\beta = 10^{-4}$	$\gamma = 10^{3}$	$\gamma = 10^{4}$	$\gamma = 10^{5}$	$\gamma_0 = 10^{3}$	$\gamma_0 = 10^{4}$	$\gamma_0 = 10^{5}$
9.6	11.0	9.6	9.5	8.7	10.5	10.8	5.3	10.4	**11.2**	8.7

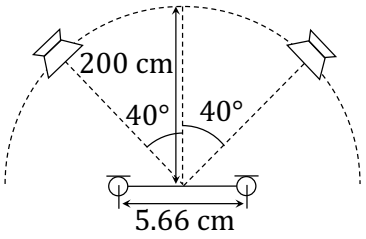
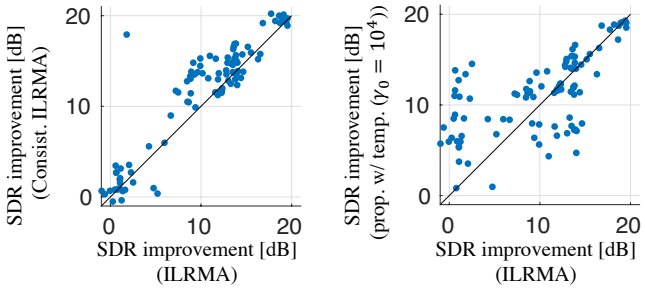


Fig. 2: Recording conditions of impulse response.



(a) (b)

Fig. 3: Pairplots for SDR improvements of ILRMA and (a) Consist. ILRMA or (b) prop. w/ temp. ($\gamma_0 = 10^4$).

the proposed method including those for the NMF variables $\{t_{ik,n}\}$ and $\{v_{kj,n}\}$ guarantee the monotonic nonincrease in the cost function (11).

C. Weight of regularization

In this paper, to prevent the effect of the regularization term from varying with the power and length of the observed signals, we normalize the regularization weight $\tilde{\gamma}$ as

$$\tilde{\gamma} = \gamma \cdot \frac{J}{\sum_{i,j} \|\boldsymbol{x}_{ij}\|_2^2} \tag{24}$$

and design $\gamma \geq 0$ instead.

In some regularization-based methods, the scheduling of the weight parameter is often performed. We use "tempering," in which the weight parameter is reduced with increasing iteration number, because the regularization to be consistent may conflict with the separation. In the tempering of γ, we set the the parameter γ at the cth iteration as $\max\{0,\ \gamma_0 - \gamma_0(c-1)/(C_0-1)\}$. Note that the parameter γ is γ_0 at the first iteration, decreases linearly thereafter, and is 0 after the C_0th iteration.

V. Simulation Experiment

A. Experimental conditions

To confirm the performance of the proposed method, a simulation experiment was conducted using two sources and two microphones ($M = N = 2$). Ten pairs of music signals from SiSEC2011[19] were prepared as dry sources. They were convolved with the impulse responses E2A ($T_{60} = 300$ ms) in the RWCP database [20]. The sources and microphones were placed as shown in Fig. 2. Both the dry sources and the impulse responses were resampled at 16 kHz. The STFT was computed with a window length of 256 ms, a hopsize of 64 ms, and the Hamming window.

We compared ILRMA [7], Consist. ILRMA [13], ILRMA with OICR (ILRMA w/ OICR) [14], the proposed method with constant γ (prop. w/o temp.), and the proposed method with the tempering of γ (prop. w/ temp.). The number of bases for NMF K was 10. The demixing matrix was initialized to the identity matrix and the NMF variables were initialized with uniformly distributed random values over $(0, 1)$. Ten trials were performed using different random seeds, i.e., we conducted experiments under 100 conditions (10 pairs of sources $\times$ 10 random seeds). The number of iterations was 400. We used $\beta = 10^{-6},\ 10^{-5},\ 10^{-4}$ in ILRMA w/ OICR, $\gamma = 10^3,\ 10^4,\ 10^5$ in prop. w/o temp., and $\gamma_0 = 10^3,\ 10^4,\ 10^5$ and $C_0 = 50$ in prop. w/ temp. The source-to-distortion ratio (SDR) [21] improvement was used as a measure of the separation performance.

B. Results and discussion

Table I shows the average of SDR improvements. Prop. w/o temp. and prop. w/ temp. improved the separation performance compared with ILRMA except for $\gamma = 10^5$ in prop. w/o temp. and $\gamma_0 = 10^5$ in prop. w/ temp., in which the initial value of γ was very large. In particular, prop. w/ temp. ($\gamma_0 = 10^4$) showed an average SDR improvement of 0.2 dB higher than Consist. ILRMA. In this experiment, ILRMA w/ OICR showed a lower separation performance than ILRMA, but this may be due to differences between the evaluation data used in [14] and in this study. Fig. 3 shows the pairplots for the SDR improvements of ILRMA with Consist. ILRMA and prop. w/ temp. ($\gamma_0 = 10^4$). Prop. w/ temp. ($\gamma_0 = 10^4$) showed higher SDR improvements than Consist. ILRMA in many cases when the SDR improvements of ILRMA were lower than 5 dB. The proposed regularization term seems to prevent the ILRMA from converging to poor solutions.

VI. Conclusion

Several methods have been proposed with the spectrogram consistency considered in ILRMA. One of them is a

regularization-based method that induces the separated signals to be consistent. Although the regularization term independent of the observed signals is used in the conventional method, we designed the regularization term that depends on the observed signals. For the optimization of the new cost function with the new regularization term, we designed the auxiliary function that is an extension of that in the conventional method. By applying VCD to the auxiliary function, we derived the update rule that guarantees the monotonic nonincrease in the cost function. A numerical experiment confirmed that the proposed method with the tempering of the weight parameter of the regularization showed a higher performance than the conventional methods.

ACKNOWLEDGMENT

The authors would like to thank Rintaro Ikeshita and Tomohiro Nakatani (NTT) for useful discussions.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.

[3] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP*, vol. 5, 2000, pp. 3140–3143.

[4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. SAP*, vol. 12, no. 5, pp. 530–538, 2004.

[5] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain bss," in *Proc. ISCAS*, 2007, pp. 3247–3250.

[6] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Proc. ICA*, 2006, pp. 165–172.

[7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

[8] D. Kitamura, N. Ono, and H. Saruwatari, "Experimental analysis of optimal window length for independent low-rank matrix analysis," in *Proc. EUSIPCO*, 2017, pp. 1170–1174.

[9] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. ASSP.*, vol. 32, no. 2, pp. 236–243, 1984.

[10] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE SP Letters*, vol. 20, no. 3, pp. 217–220, 2012.

[11] N. Perraudin, P. Balazs, and P. L. Søndergaard, "A fast Griffin–Lim algorithm," in *Proc. WASPAA*, 2013, pp. 1–4.

[12] K. Yatabe, "Consistent ICA: Determined BSS meets spectrogram consistency," *IEEE SP Letters*, vol. 27, pp. 870–874, 2020.

[13] D. Kitamura and K. Yatabe, "Consistent independent low-rank matrix analysis for determined blind source separation," *EURASIP JASP*, vol. 2020, no. 46, pp. 1–35, 2020.

[14] S. Misawa, N. Takamune, K. Yatabe, D. Kitamura, and H. Saruwatari, "Blind source separation using independent low-rank matrix analysis with spectrogram-consistency regularization," in *Proc. APSIPA*, 2023, pp. 1035–1042.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[16] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *JCGS*, vol. 9, no. 1, pp. 60–77, 2000.

[17] "Independent deeply learned matrix analysis with automatic selection of stable microphone-wise update and fast sourcewise update of demixing matrix," *Signal Processing*, vol. 178, pp. 1–12, 2021.

[18] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.

[19] S. Araki, F. Nesta, E. Vincent, *et al.*, "The 2011 signal separation evaluation campaign (SiSEC2011): - Audio source separation -," in *Proc. LVA/ICA*, 2012, pp. 414–422.

[20] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *Proc. LREC*, 2000, pp. 965–968.

[21] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.